

# 贾国庆

✉ jiaguqing12138@gmail.com · ☎ (+86) 176-3411-4423 · in Moriarty12138

## 🎓 教育背景

- 东北林业大学 (211), 哈尔滨, 黑龙江 2018 – 至今  
在读硕士研究生 软件工程, 预计 2020 年 6 月毕业
- 东北林业大学 (211), 哈尔滨, 黑龙江 2013 – 2017  
学士 软件工程

## 👨‍💻 实习经历

- 力拓飞远科技科技有限公司北京 2019 年 5 月 – 2019 年 6 月  
实习 Leader: 李嘉懿  
AI 实习生
- 文本、语音、图像数据的处理
  - 图像数据的标注
  - 文本、语音、图像比赛的 **baseline**, 包括: 中文阅读理解竞赛、细胞图像分类、城市音频分类、呼吸声音推断呼吸系统疾病、植物幼苗分类、中文的命名实体识别
  - 测测星座评论文本分类模型的建立
  - 评论文本数据的爬取
  - 微博图像数据的爬取
  - 评论分类模型的建立
- 北京魔力耳朵科技有限公司北京 2019 年 6 月 – 2019 年 8 月  
实习 Leader: 葛建军  
NLP 实习生
- 搭建 CNN-CTC、GRU-CTC、FSMN 的声学模型
  - 搭建 transformer 语言模型
  - 语音数据的处理

## 👨‍💻 项目经历

- Quora 有害问题文本分类 2018 年 12 月 – 2019 年 1 月  
PyTorch/Keras, Linux 个人项目
- 对 Quora 论坛上的问题文本进行分类. 问题文本长度较短, 且涉及有害方面较多, 包括暴力、没有建设性的问题等. 比赛要求搭建一个二分类模型从这些问题将有害问题找出来.
- 比赛地址: <https://www.kaggle.com/c/quora-insincere-questions-classification>
  - GitHub: <https://github.com/Moriarty12138/quora-insincere-question-classification>
  - 对文本数据进行预处理, 包括缩写、错误拼写等进行相应的处理
  - 采用放缩法对数据集中的数据不均衡问题进行处理
  - 对三种词向量进行平均来提高词向量的表达能力
  - 采用加权平均法集成多个模型, 包括双层 LSTM+Attention、GRU+Attention、LSTM+AVG pool+MAX pool、三层 GRU+Attention
  - 设计深度神经网络模型, 采用 k 折交叉验证进行训练
- 测测星座评论分类 2019 年 5 月 – 2019 年 6 月  
TensorFlow, Linux 公司项目

对测测星座手机软件上的评论进行分类, 涉及到的包括广告信息、人身攻击信息等. 这些评论大多特征明显. 首先考虑在传统机器学习分类算法上的分类效果, 随后采用深度神经网络来进行分类.

- 爬取评论数据, 对数据进行分析整理
- 采用 XGBoost+TF-IDF 取得较好的效果
- 搭建 CNN 模型进行训练

## 知识驱动对话系统

2019 年 3 月 – 2019 年 7 月

PyTorch, Linux 个人项目 (未完成)

目前的人机对话技术还处于初级阶段, 机器大多是被对话形式, 即机器的回复是用于响应用户的输入, 无法像人一样进行充分的对话交互, 通过设立了基于知识图谱的主动聊天任务, 让机器根据构建的知识图谱进行主动聊天, 使机器具备模拟人类用语言进行信息传递的能力。

- 建立 Encoder-Decoder 模型
- 通过知识管理器来处理上下文、知识和真实反映

## 字节跳动情感分析 + 文本点击率评估

2018 年 6 月 – 2019 年 7 月

TensorFlow, Linux 个人项目

预选赛要求建立文本情感分类模型, 用训练好的模型对测试集中的文本情感进行预测, 判断其情感为「Negative」或者「Positive」. 搜索中一个重要的任务是根据 query 和 title 预测 query 下 doc 点击率, 大赛参赛队伍需要根据脱敏后的数据预测指定 doc 的点击率.

- 预选赛比赛地址: <https://www.kesci.com/home/competition/5cb80fd312c371002b12355f>
- 正式赛比赛地址: <https://www.kesci.com/home/competition/5cc51043f71088002c5b8840>
- 搭建 LSTM 模型对文本进行情感分类
- 使用 XGBoost+TF-IDF 和 LightGBM 对文本点击率进行预测
- 搭建简单的 CNN 模型对点击率进行预测
- 分析数据, 提取特征, 提高分类准确度

## 🔧 IT 技能

---

- 编程语言: Java > Python
- Linux: 熟悉 Linux 基本操作, 有在 Linux 上的开发经验
- git: 熟悉基本 git 操作, 有在 git 上的实际开发和合作经验
- 深度学习框架: PyTorch > TensorFlow
- NLP 基础任务: 熟悉 NLP 基础, 包括分词、词性标注等
- 机器学习基础: 熟悉朴素贝叶斯、SVM 等算法
- 深度学习基础: 熟悉 RNN、LSTM、GRU、Seq2Seq 等算法
- 了解 BERT, XLNet

## ♡ 获奖情况

---

优秀奖, 中国软件杯大学生软件设计大赛  
优秀班干部  
校级三等奖学金

2016 年  
2013 年  
2014 年

## 📁 其他

---

- 技术博客: <http://www.moriarty12138.co>
- GitHub: <https://github.com/Moriarty12138>